



”Entrer dans la forteresse” : pour une édition numérique collaborative et critique de l’Encyclopédie (projet ENCCRE)

Alexandre Guilbaud, Irène Passeron, Marie Leca-Tsiomis, Olivier Ferret,
Vincent Barrellon, Sumi Yoichi

► To cite this version:

Alexandre Guilbaud, Irène Passeron, Marie Leca-Tsiomis, Olivier Ferret, Vincent Barrellon, et al..
”Entrer dans la forteresse” : pour une édition numérique collaborative et critique de l’Encyclopédie
(projet ENCCRE). Recherches sur Diderot et sur l’Encyclopédie, 2013, RDE (48), pp.225-261.
10.4000/rde.5050 . hal-01258346

HAL Id: hal-01258346

<https://hal.science/hal-01258346>

Submitted on 1 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

« Entrer dans la forteresse » : pour une édition numérique collaborative et critique de l'*Encyclopédie* (projet ENCCRE)

Le colloque « Chantiers des Lumières, l'*Encyclopédie* de Diderot et D'Alembert à l'âge de la numérisation », tenu à l'Université Paris-Diderot en mars 2013¹, visait notamment à faire le point sur la façon dont les recherches sur l'*Encyclopédie* peuvent être articulées, mises en valeur, voire facilitées par les nouvelles technologies de l'Internet et les outils informatiques développés dans ce qu'il convient aujourd'hui d'appeler le champ des Humanités numériques. Nous présentions, à cette occasion, les principes fondateurs d'un projet ambitieux et innovant, répondant à un manque flagrant : celui d'une édition numérique critique et collaborative de l'*Encyclopédie*, ou *Dictionnaire raisonné des sciences, des arts et des métiers* (1751-1772).

Ce manque que notre projet se propose de combler ne doit pas faire oublier les progrès importants réalisés ces vingt dernières années en termes d'accès au contenu de l'*Encyclopédie* et de déploiement d'outils (principalement des moteurs de recherche) permettant d'y mener un certain nombre d'investigations. Deux éditions ont clairement joué un rôle pionnier en ce sens : celle conçue par la société Redon, commercialement distribuée depuis 2000 sous forme de CD-Rom et de DVD-Rom, ainsi que celle d'ARTFL, réalisée sous la direction de Robert Morrissey (Université de Chicago), et librement accessible sur le web. La montée en puissance d'Internet au cours de la dernière décennie a progressivement conduit à l'abandon de la première, et c'est donc principalement à l'édition d'ARTFL que nous devons aujourd'hui l'avantage de pouvoir, où que l'on se trouve, lire et explorer le contenu des dix-sept volumes de textes et onze volumes de planches de l'*Encyclopédie*. Deux autres initiatives plus récentes méri-

1. <http://www.univ-paris-diderot.fr/Mediatheque/spip.php?rubrique102> (14 septembre 2013).

tent aussi d'être signalées : l'édition en ligne consultable à l'adresse <http://alembert.fr> (mais dont les éditeurs demeurent malheureusement inconnus) et la plateforme Wikisource dédiée à l'*Encyclopédie* (qui correspond avant tout à une interface collaborative proposée pour le collationnement du texte de l'œuvre, et constitue donc une réalisation numérique de nature sensiblement différente).

Pour autant, ces progrès indéniables en matière d'accessibilité et de capacité de recherche ne dispensent pas de procéder à un nouvel examen critique, pour ce qui concerne, d'une part, la qualité scientifique des éditions proposées et, d'autre part, l'articulation avec les recherches passées et actuelles sur l'*Encyclopédie*. Cet « état de l'art », dont nous rendrons compte dans un premier temps, mettra en évidence des erreurs et des lacunes qu'il nous semble important de corriger, afin que les lecteurs non spécialistes de l'ouvrage (dont une partie non négligeable d'étudiants) qui consultent chaque année ces éditions en ligne soient assurés de pouvoir accéder à une édition fiable et à la hauteur de son objet. Mais tel n'est pas le cœur de notre propos.

Au-delà de ces erreurs et de ces lacunes, nous verrons qu'aucun des sites existants n'a été conçu dans la perspective d'une connexion éditoriale forte entre l'*Encyclopédie* et les recherches qui lui sont consacrées. Ces dernières, lancées il y a plusieurs décennies par les travaux de Richard N. Schwab, Walter E. Rex, John Lough, Frank A. Kafker ou Jacques Proust², n'ont pourtant pas faibli depuis lors. La présente revue, accessible en ligne, ainsi que le précieux index qui vient récemment de la compléter, illustrent à eux seuls toute la richesse des recherches actuelles, la diversité des approches et, par là même, tout l'intérêt qu'il y aurait à valoriser l'ensemble des résultats de ces recherches au moyen d'une édition de l'*Encyclopédie* spécifiquement pensée pour les mettre en lumière, les présenter en relation étroite avec le texte de l'*Encyclopédie*, et ainsi éclairer la lecture de l'œuvre et faciliter l'appréhension de son contexte par les internautes qui seront amenés à s'intéresser à elle ou à en croiser électroniquement le chemin.

Grâce aux récentes évolutions technologiques dont le domaine des Humanités numériques bénéficie (et qu'il motive par ailleurs), ce

2. Notamment R. N. Schwab, W. E. Rex, J. Lough, *Inventory of Diderot's Encyclopédie*, SVEC 80, 83, 85, 91-93, 223, Oxford, Voltaire Foundation, 1971-1984, 7 vol. ; J. Lough, *Essays on the Encyclopédie of Diderot and D'Alembert*, London, Oxford U. Press, 1968 ; J. Proust, *Diderot et l'Encyclopédie*, 1962, rééd. Paris, Albin Michel, 1995 ; F. A. Kafker and Serena L. Kafker, *The Encyclopedists as individuals: a biographical dictionary of the authors of the Encyclopédie*, SVEC 257, Oxford, Voltaire Foundation, 1988 ; F. A. Kafker, *The encyclopedists as a group: a collective biography of the authors of the Encyclopédie*, SVEC 345, Oxford, Voltaire Foundation, 1996.

projet d'une édition numérique critique de l'*Encyclopédie* est à présent à notre portée. Il est, en somme, possible de concevoir une édition d'une autre ampleur et d'une autre nature, capable de rendre compte de l'œuvre dans toute sa complexité et d'en articuler finement les contenus, à différentes échelles, et selon différents modes de navigation, avec les résultats de la recherche. Tel est le projet, baptisé E.N.C.C.R.E. (pour Edition Numérique Collaborative et Critique de l'*Encyclopédie*), que nous souhaitons ici présenter plus en détail.

Bien entendu, la pertinence et la réussite d'une telle entreprise supposent acquis un certain nombre de préalables méthodologiques indispensables, parmi lesquels l'élaboration en amont d'une politique éditoriale rigoureuse, ainsi que l'établissement d'un dialogue étroit entre les acteurs de la recherche sur l'*Encyclopédie* et des domaines de l'ingénierie et de l'informatique appliquée aux Humanités numériques. C'est à cette condition que des innovations éditoriales et informatiques ont pu être imaginées pour relever les nombreux défis posés, d'une part, par l'œuvre elle-même, et d'autre part, par la volonté d'intégration et de mise en valeur de l'apparat critique qui structure notre démarche. Le projet ENCCRE s'appuie ainsi sur le résultat d'un travail minutieux de description des éléments caractéristiques de l'*Encyclopédie* et de l'apparat critique à éditer (ce que nous pourrions appeler les métadonnées), ainsi que des liens que ces éléments entretiennent les uns avec les autres. Il prévoit en outre le développement d'une interface collaborative d'édition spécifiquement adaptée aux pratiques de travail des chercheurs sur l'*Encyclopédie*, afin de permettre l'enrichissement critique sur le long terme et de réunir la pluralité des compétences nécessaires. Il intègre enfin d'importantes avancées dans le domaine de l'édition numérique, telles que la création de nouveaux mécanismes d'annotation, grâce auxquels tous les résultats de recherche pourront facilement trouver leur place au sein de l'édition en ligne projetée.

I. État de l'art

Le colloque « Chantiers des Lumières » des 28 et 29 mars derniers s'inscrivait dans le prolongement de deux précédents colloques internationaux consacrés à ces questions : le premier s'est déroulé en 1998 à l'EHESS sous la direction de Philippe Roger ; le deuxième, « L'*Encyclopédie* en ses nouveaux atours électroniques : vices et vertus du virtuel », a été organisé en 2000 à l'Université Paris 7 par la Société Diderot. Publiés dans le numéro 31-32 (2002) des *Recherches sur Diderot et sur l'Encyclopédie*, les actes de cette deuxième manifestation

contiennent notamment un article dans lequel Marie Leca-Tsiomis procédait à une expertise de l'exactitude des textes de l'*Encyclopédie* donnés par les deux versions électroniques de l'époque, celle en ligne d'ARTFL et celle sur CD (ou DVD) de Redon³. Elle concluait au caractère fautif, sur le plan du texte, des deux éditions, et formulait un certain nombre de recommandations de corrections « pour l'avenir » : qu'en est-il advenu ?

Délaissant l'édition de la société de Redon pour nous concentrer sur les trois éditions aujourd'hui consultables sur le web, celles d'ARTFL, de *alembert.fr* et de Wikisource, nous verrons que, si certaines de ces recommandations ont été suivies, d'autres semblent malheureusement avoir été oubliées. Outre ce premier impératif de qualité et de rigueur d'une édition qu'est l'exactitude du texte que l'on donne à lire, nous souhaiterions élargir ici notre champ d'analyse à d'autres caractéristiques de l'*Encyclopédie*, auxquelles une édition numérique doit pouvoir donner accès en termes de fonctionnalités de lecture, de navigation et de recherche. Nous débiterons donc par la question de l'exactitude du texte, avant de nous pencher plus avant sur :

— la possibilité d'accéder à la matérialité de l'ouvrage (les éditions en ligne permettent-elles de consulter et feuilleter l'*Encyclopédie* telle qu'imprimée entre 1751 et 1772 ?) ;

— la capacité à rendre compte, d'un point de vue éditorial, des modes de circulation internes à l'ouvrage (renvois entre articles au sein des volumes de textes, renvois des articles vers les planches, renvois entre planches et explications au sein des volumes de planches, renvois des volumes de planches vers les articles des volumes de textes) ;

— le respect des « désignants », c'est-à-dire des indications qui apparaissent souvent (mais pas systématiquement, loin s'en faut) à la suite du titre d'un article et qui correspondent à la branche du savoir dont relève le contenu de l'article.

Nous terminerons par l'examen de deux critères fondamentaux en matière d'édition numérique : d'abord l'efficacité du moteur de recherche et la pertinence des champs de recherche proposés, puis la question des métadonnées et de leur structuration, c'est-à-dire la façon et le degré de finesse avec lesquels les éditions décrivent et articulent informatiquement les éléments constitutifs (articles, planches, explications de planches) de l'*Encyclopédie*.

3. M. Leca-Tsiomis, « Numérisations et exactitude du texte encyclopédique : quelques propositions pour l'avenir », *RDE* 31-32, 2002, p. 293-302.

L'exactitude du texte

Donner à lire un texte fidèle à l'original selon des normes éditoriales explicites : tel doit être le premier dessein de toute édition (numérique ou non) et de tout éditeur. Dans le cas de l'*Encyclopédie*, le défi est d'autant plus ardu que l'ouvrage renferme une quantité non seulement très importante de texte (17 volumes de mille pages in-folio chacun environ, auxquels s'ajoutent notamment les explications des volumes de planches), mais aussi très riche, que ce soit par la présence d'une large gamme de caractéristiques (et de variantes) typographiques, de nombreuses langues étrangères, de formules mathématiques, ou d'une multitude de tables et de tableaux.

Partons du plus simple : la qualité de la saisie (et donc du collationnement) du texte brut. D'importants progrès dans ce domaine ont été réalisés par l'édition d'ARTFL depuis le colloque de l'an 2000. Comme le signale Robert Morrissey dans son « Introduction à l'*Encyclopédie* d'ARTFL »⁴, plusieurs phases importantes de correction⁵ ont été conduites entre 1999 et 2008. Néanmoins, les nombreux sondages que notre équipe a pu effectuer montrent que la saisie du mode texte demeure encore imparfaite. Le texte de l'édition *alembert.fr* nous a, quant à lui, paru d'une qualité très légèrement supérieure pour ce qui est du collationnement, mais souvent en deçà du point de vue du rendu typographique (les italiques y sont par exemple nettement moins bien restitués que sur ARTFL).

Pour le reste, le bilan est en demi-teinte. Aucune de ces deux éditions ne donne accès aux *errata* publiés par les éditeurs de l'*Encyclopédie* dans son corps même. Le riche langage mathématique sur lequel s'appuient de nombreux articles scientifiques (et quelques articles d'art) n'est pas transcrit, dans un cas comme dans l'autre. À l'exception du grec pour ARTFL, les caractères d'écriture non romains ne sont pas restitués. Quant au découpage alphabétique original – dans lequel I et J, d'une part, et U et V, d'autre part, sont confondus –, il n'est pas respecté⁶.

4. Voir <http://encyclopedia.uchicago.edu/node/161> (29 juin 2013).

5. Voir <http://encyclopedia.uchicago.edu/node/13> (29 juin 2013).

6. Contrairement à ce que l'on pourrait croire en dissimulant I et J, l'article JÉSUISTE est bien dans le tome VIII qui va de H à IT. Ledit article étant bien daté de 1762, on voit que ce fameux tome VIII n'était pas terminé au moment de l'interdiction de 1759.

spécifier certains éléments caractéristiques de l’*Encyclopédie*, tels que les adresses, les renvois entre articles – rendus sous la forme de liens directs vers les adresses « cibles » –, ainsi que les changements de page et de colonne.

Mais dans la pratique, il faut agir ici comme dans le premier problème ; vous n’aurez seulement qu’à diviser 1411200 par la longueur, vous aurez le carré du nombre des vibrations ; de même que l’on divise ce nombre par le carré des vibrations pour trouver la longueur.

Sur ces principes M. Derham a construit une table des vibrations des *pendules* des différentes longueurs dans l’espace d’une minute.

Longueur du pendule en pouces.	Vibrations en une minute.	Longueur du pendule en pouces.	Vibrations en une minute.
1.	375.7.	30.	68.6
2.	265.6.	39.2.	68.0
3.	216.9.		
4.	187.8.	40.	59.5.
5.	168.0		
6.	153.3.		
7.	142.0.	50.	53.1.
8.	132.8.	60.	48.5.
9.	125.2.	70.	44.9.
10.	118.8.	80.	42.0.
20.	84.0.	90.	39.6.
		100.	37.5.

Remarquez que ces lois du mouvement des *pendules* ne s’observeront pas à la rigueur, à moins que le fil qui soutient la boule n’ait aucun poids, & que la pesanteur de tout le poids ne soit réunie en un seul point.

Méthode générale pour trouver le mouvement d’un pendule. Soit *a* le rayon du cercle que décrit le *pendule*, ou la longueur du *pendule* ; *b*, l’abscisse totale qui répond à l’arc du centre, en prenant cette abscisse depuis le point le plus bas ; *x*, l’abscisse d’une portion quelconque de cet arc ; *p*, la pesanteur ; *u*, la vitesse en un point quelconque, on aura $uu=2p(b-x)$. Voyez les articles FORCE ACCÉLÉRATRICE & PLAN INCLINÉ) ; & le tems employé à parcourir un arc quelconque infiniment petit, sera $\frac{-adx}{a\sqrt{2ax-xx}} = \frac{-adx}{\sqrt{2ax-xx}} \times \frac{1}{\sqrt{2p}\cdot\sqrt{b-x}}$. Or, lorsque l’arc descendu n’a pas beaucoup d’amplitude, *x* est petit par rapport à *a* ; & on peut, au lieu de $\frac{1}{\sqrt{2ax-xx}}$, ou $\frac{1}{\sqrt{x+2a-x}}$ écrire $\frac{1}{\sqrt{x}} \times \left(\frac{1}{\sqrt{2a}} + \frac{x}{4a\sqrt{2a}} \right)$, &c. (voyez BINOME, APPROXIMATION, & EXPOSANT) ; de manière que l’élément du tems sera à-peu-près $\frac{-a}{\sqrt{1p}} \times \left(\frac{dx}{\sqrt{2a}\sqrt{2x-xx}} + \frac{xdx}{4a\sqrt{2a}\sqrt{bx-xx}} \right)$, &c. quantité qui étant intégrée par les règles connues, donnera à-peu-près le tems d’une demi-vibration du *pendule*. On peut même, lorsque l’arc descendu est fort petit, négliger entièrement le terme $\frac{+xdx}{2a\sqrt{2a}\sqrt{bx-xx}}$; & alors le tems de la descente du *pendule* sera sensiblement le même que celui de la descente dans une cycloïde qui auroit le rayon osculateur à son sommet égal au rayon du *pendule*.

Illustration 2. Copie d’écran de deux extraits de l’article PENDULE édités sur Wikisource (*Enc.*, XII, p. 297a et 297b respectivement)

Les résultats sont là : à partir de plusieurs sources initiales (dont ARTFL et alembert.fr) complétées par les outils d’océrisation disponibles à la demande sur l’interface, la saisie collaborative de l’*Encyclopédie* sur Wikisource est aujourd’hui devenue la meilleure collation disponible sur le web. Elle reste certes encore partielle (plusieurs tranches ne sont pas encore intégralement corrigées, et certains volumes de planches font actuellement défaut), mais la vitesse avec laquelle le collationnement avance montre qu’il sera très certainement achevé d’ici un ou deux ans, et nous incite à y apporter nos propres contributions. Notons d’ailleurs que les données générées dans ce cadre sont disponibles sous licence « Creative Commons paternité – partage à

l'identique des conditions initiales »⁹, ce qui autorise à les utiliser librement, sous condition d'en publier le résultat sous une licence identique ou équivalente.

Au-delà du présent état de l'art, les progrès réalisés grâce au Wikisource montrent surtout à quel point certaines évolutions technologiques récentes conduisent à envisager autrement l'édition d'une œuvre aussi monumentale et complexe que l'*Encyclopédie*. Le modèle de travail collaboratif grâce à une plateforme d'édition pensée *ad hoc* joue, nous le verrons, un rôle fondamental dans le projet ENCCRE.

Accès à la matérialité de l'ouvrage

Notre deuxième critère d'analyse concerne la capacité de l'édition numérique à fournir les moyens d'accéder à ce qui constitue la matérialité de l'ouvrage¹⁰. Il s'agit, d'une part, pour le lecteur, de pouvoir se reporter à tout moment au fac-similé original correspondant à l'article ou à la planche (ou son explication) qu'il est en train de consulter et, d'autre part, de pouvoir « feuilleter » et se repérer dans un volume de textes ou de planches de l'*Encyclopédie* tel qu'il a été imprimé à l'époque, comme s'il l'avait (numériquement) entre les mains.

L'édition *alembert.fr* ne répond à aucune de ces deux exigences. On accède à un article *via* un unique classement alphabétique (présenté sous la forme d'une liste de regroupements de lettres cliquables surplombant l'espace de lecture du site), sans aucun moyen de savoir dans quel tome ni à quelles pages ledit article a été publié. Les fac-similés des volumes de textes sont introuvables. Quant aux volumes de planches, on pourra certes les consulter à loisir sur une interface conviviale et bien pensée pour la visualisation d'images, mais sans pouvoir se repérer dans l'édition, ni même avoir accès aux explications qui les accompagnent pourtant thème par thème dans les volumes originaux de l'*Encyclopédie*.

A contrario, l'édition d'ARTFL a réalisé des progrès notables dans ce domaine depuis le colloque de l'an 2000. Les fac-similés des originaux (volumes de textes et volumes de planches) sont disponibles et l'interface de lecture a très récemment été dotée d'une nouvelle fonctionnalité « browse »¹¹ grâce à laquelle le lecteur peut lire les

9. <http://creativecommons.org/licenses/by-sa/3.0/deed.fr> (29 juin 2013).

10. Notons qu'il ne s'agit en aucun cas de restreindre l'édition numérique à cette seule fonctionnalité de lecture, mais bien plutôt d'offrir systématiquement cette fonctionnalité comme l'un des modes de consultation numérique *possibles* de l'*Encyclopédie*.

11. <http://encyclopedia.uchicago.edu/content/browse> (29 juin 2013).

articles ou les explications et les planches associées selon leur ordre original d'impression. Quelques erreurs subsistent néanmoins¹² et, de façon plus générale, les informations de localisation des articles consultés (tome, pagination et indication de colonne) restent toujours difficilement accessibles.

En permettant la mise en regard de chaque page de l'*Encyclopédie* avec le fac-similé original correspondant, que ce soit pour la lecture ou la correction du texte, l'interface Wikisource offre quant à elle le rendu éditorial le plus fidèle de ce point de vue (voir Illustration 1). D'autres modes de parcours permettent en outre d'explorer l'œuvre volume par volume, ou de la feuilleter page par page, au plus près de l'édition originale. Restent cependant quelques erreurs à corriger ainsi qu'un important travail de collationnement à effectuer sur les volumes de planches manquants pour pouvoir disposer de cette importante fonctionnalité à l'échelle de l'ensemble de l'ouvrage.

Les modes de circulation internes à l'ouvrage

Pour les raisons bien connues et longuement développées par Diderot et D'Alembert dans le *Prospectus* de 1750, le « Discours préliminaire » paru dans le tome I (1751) ou les articles DICTIONNAIRE (*Enc.*, IV, p. 958b-970a) et ENCYCLOPÉDIE (*Enc.*, V, p. 635a-648b), l'*Encyclopédie* concilie deux logiques de lecture dans un même ensemble : celle d'un dictionnaire, *via* le classement par ordre alphabétique, et celle de l'ordre encyclopédique permettant de faire apparaître « l'ordre et l'enchaînement des connoissances humaines » (*Enc.*, I, p. i). Ce dernier est notamment réalisé au moyen de nombreux renvois internes à l'ouvrage : renvois entre articles, complétés par les renvois des articles vers les planches, des volumes de planches vers les articles des volumes de textes, sans oublier les liens entre les planches et leurs explications. Nous avons rassemblé dans le tableau suivant les résultats de notre analyse sur la capacité des trois éditions à rendre compte de ces différents modes de circulation. Les lacunes sont évidentes pour ce qui concerne la restitution des liens impliquant les volumes de planches. Les renvois entre articles méritent, quant à eux, quelques éclaircissements complémentaires.

12. D'après la page « browse », le premier tome commence apparemment par le « Discours préliminaire » de D'Alembert, et il en est de même pour les « Avertissements » dans les sept tomes suivants. Recherches faites, les pages de titres et épîtres de chaque volume sont en fait accessibles *via* la deuxième partie cliquable de l'intitulé du volume (exemple : « A - Azymites, June 1751 » pour le tome I).

	alembert.fr	ARTFL	Wikisource
Renvois entre articles de texte	par le biais du moteur de recherche	par le biais du moteur de recherche	renvoi direct vers l'adresse
Renvois des articles vers les planches	non	non	non
Renvois des volumes de planches vers les volumes de textes	non	non	non
Liens entre une planche et son explication	non	oui	non

La consultation des éditions ARTFL et alembert.fr montre en effet que les renvois entre articles n'y sont pas rendus sous la forme d'un lien direct entre le renvoi et l'article pointé, mais par le moyen indirect du moteur de recherche. Dans le cas d'ARTFL (il en est de même sur alembert.fr), le clic sur un renvoi au sein d'un article ne redirige donc généralement pas seulement vers l'article effectivement concerné par le renvoi, mais à la liste des adresses et entrées de l'*Encyclopédie* dont l'intitulé contient le terme (ou le groupe de termes) constituant ce renvoi. Le renvoi « TERRE » de l'entrée FIGURE DE LA TERRE (*Enc.*, VI, p. 749b), par exemple, conduit ainsi à une page donnant la liste des 208 articles dont le titre contient le mot « Terre » – ce qui inclut par exemple PATATE, TOPINAMBOUR ou POMME DE TERRE, TERREAU, ou encore PEUR, FRAYEUR, TERREUR. Difficile, dans ces conditions, de retrouver l'article vers lequel D'Alembert souhaitait précisément renvoyer le lecteur à l'intérieur de l'ouvrage.

L'interface Wikisource permet au contraire de déclarer ces mêmes renvois sous forme de liens directs vers un article « cible » dans l'ouvrage. Quoique conforme à la logique éditoriale des renvois dans l'*Encyclopédie*, ce dispositif se heurte néanmoins à deux difficultés. Il se trouve premièrement limité par l'impossibilité de spécifier les différentes entrées d'une même adresse¹³ dans l'interface de collationnement proposée : un renvoi ne « ciblera » qu'une adresse, sans pouvoir nous conduire plus précisément vers l'entrée visée – si tant est que cette information soit donnée par l'auteur du renvoi. En somme, le disposi-

13. Nous utilisons ici la terminologie qui permet de distinguer l'entrée générale TERRE (en général en grandes capitales) que nous appelons « adresse », des sous-entrées TERRE, *en Astronomie* (en général en petites capitales) que nous appelons « entrées ». Une adresse peut avoir plusieurs entrées.

tif requiert donc une description de la structure des articles plus fine que celle actuellement proposée par l'interface Wikisource. La saisie précise d'un renvoi nécessite aussi de pouvoir identifier avec certitude l'adresse (ou l'entrée) visée par l'auteur, ce qui, comme on sait, pose fréquemment problème, soit parce qu'il demeure un doute sur l'article en question, soit parce que l'article annoncé dans le renvoi n'existe pas, diffère plus ou moins sensiblement du titre de l'article existant, etc. Si l'implémentation des renvois sous forme de liens directs paraît la solution éditoriale la plus naturelle et la plus adaptée, elle ne suffit cependant pas toujours et devra donc parfois être couplée à des informations critiques permettant de justifier les choix effectués (sans introduire de confusion avec ce qui existe factuellement dans l'œuvre originale).

L'édition des renvois sur le site ARTFL soulève en outre une autre difficulté, déjà soulignée par Marie Leca-Tsiomis¹⁴ dans les actes du colloque de l'an 2000 : les renvois y sont automatiquement identifiés sur la base des seules caractéristiques typographiques des portions de textes concernées. En d'autres termes, un mot ou groupe de mots sera systématiquement considéré comme un renvoi dès lors qu'il apparaît en petites capitales et qu'il ne correspond pas à une entrée. C'est oublier que les petites capitales sont aussi fréquemment utilisées dans l'*Encyclopédie* afin de marquer certaines des signatures d'articles ainsi que certains des noms propres cités à l'intérieur d'un article. C'est aussi oublier que les renvois apparaissent régulièrement sous d'autres formes moins explicites¹⁵. Cette situation conduit inévitablement à des lacunes ainsi qu'à des erreurs manifestes. Celles relatives à la recherche des petites capitales sont à présent plus faciles à repérer, grâce à l'outil expérimental¹⁶ que l'équipe ARTFL vient récemment de mettre à disposition et qui permet de rechercher des termes parmi les renvois identifiés dans les dix-sept volumes de texte. La recherche du terme « Jaucourt » donne par exemple une liste de 22 articles dans lesquels, vérifications faites, la signature (de Jaucourt) se trouve invariablement assimilée à un renvoi (en raison de la marque des petites capitales). Nous aboutissons à des résultats de même nature avec Boulanger, de Ratte, Rousseau, ou Tronchin (la liste n'est évidemment pas exhaustive), ainsi qu'à une confusion systématique entre renvois et noms de collaborateurs dans la liste de « Marque des auteurs » disposée sur la

14. Voir art. cité, note 3, p. 295.

15. Tel est le cas, parmi tant d'autres, du renvoi suivant aux articles GERBE et BOUILLON, donné par d'Argenville dans l'article *JET d'eau* : « Il y a encore des gerbes, des bouillons. Consultez *ces articles* à leur lettre » (*Enc.*, VIII, p. 521a).

16. <http://encyclopedie.uchicago.edu/content/renvois-navigation> (30 juin 2013).

dernière page d'articles du tome III (p. 905). Espérons donc que ce moteur de recherche expérimental puisse faciliter la correction de ces multiples ratés.

Quoi qu'il en soit, l'édition du système de renvois par le biais de méthodes automatiques (utilisation du moteur de recherche pour la navigation à partir du renvoi, repérage des renvois grâce aux petites capitales) montre clairement ses limites, en se révélant finalement incapable de restituer l'un des principaux modes de circulation prévus par les deux éditeurs. Quant à la solution consistant à implémenter les renvois sous forme de liens vers un article « cible », elle nécessite d'abord une description précise du découpage des articles (adresses, entrées, etc.) et, dans un second temps, la capacité d'introduire des informations critiques permettant d'éditer en toute rigueur les cas de figure les plus délicats.

La question des « désignants »

Après le critère des renvois, penchons-nous sur les deux autres principaux moyens de mise en ordre encyclopédique de l'ouvrage : le « Système figuré des connoissances humaines » et les désignants. Le premier, non accessible sur *alembert.fr*, est fidèlement reproduit sur ARTFL et Wikisource. Le second pose quant à lui différents types de problèmes.

Rappelons d'une part que les désignants constituent l'un des moyens, avec le « Système figuré », mais de façon distincte, de rattacher les articles à une « matière ». Ils apparaissent souvent entre parenthèses et en italique après la vedette d'entrée, mais aussi sous de nombreuses autres formes : il peut s'agir d'une typographie différente, de désignants multiples, de désignants chaînés ou de désignants absents mais rendus explicites quelque part dans l'article¹⁷. Rappelons d'autre part que toutes les entrées ne portent pas de désignants, soit parce que certaines relèvent d'une trop grande variété d'arts, de métiers, de sciences ou de matières pour pouvoir être rattachées à un domaine en particulier, soit parce que le collaborateur chargé de l'article n'a pas jugé nécessaire ou a oublié d'en indiquer un.

17. D'Alembert a prévenu le lecteur dans le « Discours préliminaire » : « S'il arrive que le nom de la Science soit omis dans l'article, la lecture suffira pour connoître à quelle Science il se rapporte ; & quand nous aurions, par exemple, oublié d'avertir que le mot *Bombe* appartient à l'art militaire, & le nom d'une ville ou d'un pays à la Géographie, nous comptons assez sur l'intelligence de nos lecteurs, pour espérer qu'ils ne seroient pas choqués d'une pareille omission. » (*Enc.*, I, p. xvij).

Depuis plusieurs années, l'édition d'ARTFL nous permet de repérer un certain nombre de ces désignants. Beaucoup d'entre eux, d'après les sondages effectués, ont cependant été oubliés, soit parce que le découpage entre l'entrée et le désignant correspondant a été mal réalisé, soit parce que le désignant se trouve trop loin de l'entrée, ou typographié autrement qu'en italique, pour pouvoir être identifié. Alain Cernuschi faisait ainsi remarquer en 2002 que 60 % des articles musicaux munis de désignants n'étaient pas reconnus¹⁸. À ce niveau encore, un travail méticuleux de relecture s'impose.

L'édition d'ARTFL offre aussi le moyen de rechercher dans la « classification »¹⁹ de l'*Encyclopédie* (comprendre « parmi ses désignants »), ainsi que dans une forme normalisée française et anglaise (accessible *via* le moteur de recherche avancée) de cette « classification » – les éditeurs ont ici eu recours à une « normalisation » conduite suivant des règles non explicitées, et distincte, en tous cas, des abréviations utilisées par Richard N. Schwab dans son *Inventory of the Encyclopédie*²⁰. Outre le fait qu'une recherche sur un ensemble d'éléments insuffisamment renseignés risque de conduire à des résultats non représentatifs – et donc nécessairement erronés –, le terme de « classification » pose ici un problème crucial : faire croire au lecteur (mal informé par les deux formulaires de recherche, simple et avancée) que le système des désignants (originaux ou « normalisés ») de l'*Encyclopédie* puisse correspondre à une véritable classification des savoirs est en effet une erreur, scientifiquement parlant, pour plusieurs raisons.

Il pourrait d'abord être tentant d'envisager le « Système figuré » comme un moyen de classer les articles en fonction de ses branches et ramifications – et de légitimer ainsi l'existence d'une telle classification au sein de l'*Encyclopédie*. Un rapide examen permet cependant de constater que chaque article ou presque relève de plusieurs régions du « Système figuré », et que cet arbre des connaissances ne correspond

18. A. Cernuschi, « La question d'un découpage par matières : l'exemple du corpus musicographique de l'*Encyclopédie* », *RDE* 31-32, 2002, p. 165. En reprenant la comparaison le 30 juin 2013, on voit immédiatement qu'au regard des 1168 articles du corpus musical de l'*Encyclopédie* (A. Cernuschi, *Penser la musique dans l'Encyclopédie*, Paris, Champion, 2000, p. 673-696, et *ibid.*, p. 167), les 462 résultats obtenus en interrogeant ARTFL par la « classification » « musique » demeurent très loin du compte. Une des raisons en est toujours l'insuffisante reconnaissance des désignants, à laquelle s'ajoutent les problèmes explicités dans ce qui suit.

19. Le terme anglais utilisé sur le site est le même : voir <http://encyclopedia.uchicago.edu/node/176> (30 juin 2013). Les termes anglais originaux apparaissent par la suite entre parenthèses, à la suite de leur équivalent français.

20. *Inventory* (ouvr. cité, note 2), t. II, p. 12-15.

donc pas à une partition, mais bien plutôt à une cartographie possible parmi d'autres²¹. Des chercheurs de spécialités très diverses se sont d'ailleurs réunis en 2004 pour confronter leurs travaux sur la complexité des désignants. De la richesse de ces études²² s'est dégagée une certitude : rien de systématique ni de parfaitement cohérent ne peut être observé dans cet ensemble de désignations. Les multiples termes utilisés à cet effet sont totalement disparates dans leur statut : parfois multiples, parfois chaînés, parfois tout simplement absents, ces désignants se situent à tous niveaux de précision, sans suivre de logique hiérarchique ou classificatoire particulière. Ils se recouvrent, se contredisent parfois, ou concernent même tout autre chose qu'un domaine du savoir ou une « matière » (nous pensons par exemple aux désignants de Jaucourt qui multiplie les termes s'apparentant à des parties de définitions²³).

L'équipe d'ARTFL a néanmoins poussé plus loin une logique fondée sur une hypothétique organisation des désignants en concevant un outil de « classification par la machine » (« machine class ») des articles de l'*Encyclopédie*. Partant du constat selon lequel 15 000 articles environ restent sans désignant dans leur édition numérique de l'ouvrage, et qu'un total de près de 3 000 désignants normalisés différents (plus de 8 500 en considérant les désignants originaux repérés par ARTFL) constituerait un ensemble trop large pour être en mesure de fournir une classification utilisable, les éditeurs se sont lancés dans une expérience visant à restreindre la liste et à attribuer un désignant à tous les articles, grâce à un outil informatique permettant d'analyser ces articles comme des « sacs de mots » et d'en détecter quantitativement la proximité thématique. Au total, la « classification par la machine » retient 360 classes parmi les 3 000 normalisées (sans aucune explicitation des critères de choix des classes retenues) et ne laisse plus apparaître que 8 articles non classifiés (« unclassified » ou « pending »). L'examen de ces 360 « classes »²⁴ et de la répartition des quelque 74 000 articles est édifiant.

21. Voir notamment l'exemple des branches « mathématiques mixtes » et « sciences physico-mathématiques » étudié par Irène Passeron : « D'Alembert refait le MONDE (*Phys.*) : parcours dans les mathématiques mixtes », *RDE* 40-41, 2006, p. 155-177.

22. Elles sont publiées dans le numéro 40-41 (2006) des *Recherches sur Diderot et sur l'Encyclopédie*.

23. Ainsi MOUCHE GRUE (*Science microsc.*), TRÉTEAU (*instrument d'Ouvrier*), VAMPIRE (*Hist. des superst.*), VÉLIE (*Topogr. de Rome*).

24. Selon le classement par la machine, « Mathématique » ne contient, étonnamment, que 10 articles ! « Metteur en œuvre » en contient 89 ; « Ustensile de ménage », 1.

Les résultats auxquels l'édition d'ARTFL est ainsi parvenue conduisent à remettre en question non seulement l'utilité d'une telle expérience, mais aussi la rigueur de la méthode selon laquelle l'attribution informatique a été effectuée – méthode bien opaque, et d'ailleurs apparemment complétée par un certain nombre d'ajustements tout aussi obscurs dans leurs principes, si l'on s'en réfère aux explications données sur le site. Les très nombreux sondages que nous avons réalisés (et dont nous ne donnons ici que quelques exemples) sont révélateurs : le classement (par la machine) de l'entrée mathématique ABAQUE, ou Table de Pythagore, en « manufacture de glaces » ; celui de l'article, tout autant mathématique, ANALYSE (possédant le désignant chaîné original « *Ordre encyclop. Entend. Raison. Philosoph. ou Science, Science de la Nature, Mathématiques pures, Arithmétique littérale, ou Algèbre, Analyse* ») de D'Alembert en « chimie » ; le classement de POU, LE, que l'*Encyclopédie* désigne sans ambiguïté comme un terme d'Astronomie chinoise, en « histoire naturelle » ; ou encore celui de l'article ROSE-CROIX, *société des frères de la*, en « histoire de la philosophie », alors que Jaucourt avait pourtant pris la peine d'inventer le cinglant désignant « *histoire des impostures humaines* ». C'est avec un certain soulagement que nous saluons donc la récente décision de l'équipe d'ARTFL de faire disparaître son moteur de recherche sur les « machine class » de son formulaire de recherche avancée. Quoiqu'il n'ait pas été définitivement supprimé – il est désormais proposé parmi les outils de navigation (« Navigation Tools ») accessible *via* le menu de gauche du site²⁵ –, cet outil répandra toutefois moins d'erreurs dans l'esprit des lecteurs à cette nouvelle place qu'à l'ancienne.

Signalons pour finir que l'édition d'ARTFL ne constitue pas un cas isolé. L'édition alembert.fr, dans laquelle le travail de repérage semble tout juste amorcé, commet la même confusion en assimilant peut-être plus abruptement encore les désignants de l'ouvrage à un système classificatoire partiel. Les désignants sont regroupés sans explication sous la bannière de neuf domaines plus ou moins tirés du « Système figuré » (« Arts & métiers / Histoire / Histoire naturelle / Religion / Logique / Morale / Science / Physique particulière / Beaux-arts »), présentée au centre de la page d'accueil du site comme un moyen de navigation, et complétée par un dernier item rassemblant le « Non classifié »... Pour ce qui est de Wikisource, aucune fonctionnalité d'édition ne permet actuellement d'y spécifier les désignants, que ce soit de façon manuelle ou automatique.

25. <http://encyclopedia.uchicago.edu/content/machine-classifications> (30 juin 2013).

Ces choix éditoriaux fort contestables témoignent d'une connaissance par trop fragile des recherches sur l'*Encyclopédie*. La lecture de la présente revue, parmi d'autres travaux, nous rappelle en effet que la disparité des désignants constitue l'une des plus grandes richesses de l'ouvrage, en permettant d'y lire et d'y déceler tout le foisonnement intellectuel de l'entreprise, les soubresauts de son histoire et la diversité de ses collaborateurs, de leur mode opératoire, de saisir sur le vif la confrontation de découpages traditionnels et l'apparition de nouvelles sciences ou méthodes (la science microscopique, la nouvelle chimie). Ici encore, la forte hétérogénéité des éléments constitutifs de l'*Encyclopédie* interdit *de facto* d'espérer parvenir à une édition fiable et rigoureuse par le seul moyen d'un traitement automatisé des données.

Moteur de recherche et structuration des données

Le moteur de recherche forme comme on sait l'une des fonctionnalités centrales d'une édition numérique : celle permettant d'effectuer des recherches sur le corpus édité par le biais de requêtes. Au-delà de son efficacité, la qualité d'un tel outil repose néanmoins sur deux autres prérequis fondamentaux : la fiabilité des données sur lesquelles le moteur effectue les requêtes, et la pertinence, compte tenu du corpus concerné, des différents champs pouvant être interrogés par le moteur. La question est, en cela, non seulement étroitement liée à la qualité de collationnement du texte, mais aussi au choix des éléments caractéristiques qui ont été spécifiés au sein de ce texte (des métadonnées, donc, que nous qualifierons d'*encyclopédiques* dans le sens où elles sont propres à cet ouvrage), ainsi qu'à la façon et la rigueur avec lesquelles ces éléments ont été repérés. Un moteur de recherche, par exemple, ne permettra pas d'effectuer une enquête sur les titres des articles (adresse et/ou entrées) de l'*Encyclopédie* si ces derniers n'ont pas été systématiquement déclarés comme tels, et il ne le permettra pas efficacement si le travail de repérage et de spécification correspondant n'a pas été effectué avec toute la rigueur nécessaire. L'établissement de la liste de ces éléments caractéristiques (adresse, entrée, renvois, signatures, etc.) et l'identification du type de relations qui les lient entre eux (tel renvoi appartient à tel article, par exemple) relèvent de ce que nous appellerons la *structure des données*. La méthode informatique selon laquelle ces éléments et les liens correspondants sont déclarés correspond au travail d'*encodage* des données.

Des trois éditions, celle d'ARTFL possède de loin le moteur de recherche le plus puissant et le mieux conçu. L'outil proposé sur le site alembert.fr se limite en effet à l'implémentation de la version « customisée » du moteur de recherche de Google (choix rudimentaire, peu

adapté à l'*Encyclopédie*). Quant à l'interface Wikisource, prioritairement dédiée au collationnement, elle n'offre pas la possibilité d'effectuer des requêtes directes sur le texte de l'œuvre.

Le formulaire de recherche simple d'ARTFL²⁶ permet au contraire d'interroger l'ensemble du contenu de l'ouvrage en mode plein-texte, de rechercher dans la liste des adresses et des entrées, des noms des auteurs des articles, ainsi que dans ce qu'ARTFL appelle la « classification » de l'*Encyclopédie* (c'est-à-dire les désignants). L'outil de recherche avancée²⁷ étend quant à lui les possibilités de raffinement de la requête (recherche de termes séparés dans le texte par un nombre donné de mots, etc.) et élargit notamment les champs interrogeables à la version normalisée (en français et en anglais) des désignants et au type de contenu (volume de texte ou volume de planches). L'ensemble reste bien sûr fortement tributaire du niveau d'exactitude du texte et du repérage des renvois, sur lesquels, nous l'avons déjà noté, des efforts sérieux méritent encore d'être réalisés. Néanmoins, la limitation la plus contraignante du moteur de recherche d'ARTFL et, plus généralement, de l'édition proposée, tient à la structure de données sur laquelle elle s'appuie et, partant, aux choix de spécification qui ont été retenus. La consultation du « code » caché derrière la page d'affichage de n'importe quel article montre en effet que les métadonnées associées (adresse ou entrée, type d'objet, auteur, « classe », classe normalisée, classe normalisée anglaise, classe fournie par la machine : nous retrouvons logiquement tous les champs des formulaires de recherches simple et avancée) sont toutes déclarées extérieurement au texte qui le compose : elles s'y appliquent donc globalement, indifféremment et inconditionnellement. La seule exception à cette règle n'en est pas vraiment une : il s'agit des renvois qui, à défaut d'être spécifiés comme tels (et donc décrits comme métadonnées) dans le texte, le sont sur la seule base d'un critère typographique, celui des petites capitales. Cette manière de procéder limite intrinsèquement les possibilités de faire évoluer l'édition, sa compréhensibilité et les moyens d'interrogation qui lui sont associés.

Prenons l'exemple de la métadonnée « auteur ». Les noms des auteurs, comme on sait, apparaissent très rarement de façon explicite dans l'*Encyclopédie*. Les articles contiennent au contraire des signatures – un (O) pour D'Alembert, une étoile « * » pour Diderot, etc. – par le biais desquelles il sera possible d'identifier le collaborateur correspondant pour peu que l'on sache à quels endroits de l'ouvrage se

26. <http://encyclopedia.uchicago.edu/node/176> (30 juin 2013).

27. <http://encyclopedia.uchicago.edu/node/175> (30 juin 2013).

trouvent les tables de correspondance entre signatures et auteurs. Comment donner accès à cette information si ces signatures n'ont pas été spécifiées comme telles dans l'édition, et liées au nom des auteurs auxquelles elles réfèrent ? Comment traiter le cas des nombreux articles signés par deux collaborateurs ou plus, et indiquer au lecteur les passages de l'article qui peuvent respectivement leur être attribués lorsque les noms des collaborateurs sont, comme sur ARTFL, tous indistinctement associés à l'ensemble du texte de l'article ? Comme tant d'autres dans l'*Encyclopédie*, la question de l'attribution des articles conduit à prendre en compte une multitude de situations particulières nécessitant la mise au point d'une structure de données spécifiquement adaptée, à même de les rendre intelligibles aux yeux du lecteur en même temps que d'offrir des fonctionnalités de recherche permettant de conduire des requêtes précises, pertinentes, et susceptibles de mener à de nouvelles découvertes.

De la même façon, nous avons pu constater que l'impossibilité de spécifier les entrées *en tant que telles* sur Wikisource grève intrinsèquement le système de description des renvois (dès lors restreint à des liens vers les seules adresses). Quoique cette interface ne possède pas de moteur de recherche proprement dit, et n'ait donc pas actuellement pour fonction de fournir de tels moyens d'interrogation, les choix effectués en terme de structuration des données n'en ont pas moins des conséquences directes sur les fonctionnalités de navigation de l'édition proposée.

C'est, nous allons le voir, en répondant à cette exigence primordiale de structuration des données que nous sommes en mesure de proposer un autre projet d'édition, fondamentalement différent dans sa nature des réalisations existantes : celui d'une édition numérique *critique*, fidèle à son objet, et enrichie par les résultats des nombreuses recherches, passées, actuelles et à venir sur l'*Encyclopédie*.

Ce projet repose plus généralement sur une conception différente du rôle et de l'usage des outils informatiques dans le domaine de l'édition numérique. Nous avons pu précédemment constater que nombre des erreurs et lacunes relevées dans les éditions existantes trahissent une connaissance souvent insuffisante de l'*Encyclopédie* et de l'histoire concrète de son élaboration. Dans la plupart des cas, la spécification de ses éléments caractéristiques a non seulement été effectuée sur la base de principes inadaptés à ce qu'est l'*Encyclopédie* et ce que nous savons à son sujet, mais elle l'a également été grâce à des procédures informatiques conduites à grande échelle, sans vérification des résultats obtenus (c'est-à-dire en *lisant*, tout simplement, le

texte)²⁸. Si les logiciels d'analyse, de repérage ou de comparaison automatiques de données constituent indiscutablement des outils utiles, pour ne pas dire indispensables, à l'édition numérique de corpus de grande taille, leur utilisation doit cependant répondre à certaines contraintes, spécifiques de l'œuvre éditée, et garantes, en cela, d'un niveau minimum de qualité et de rigueur éditoriale. Il s'agit, en d'autres termes, d'articuler l'approche quantitative, rendue possible par les nouvelles technologies, avec l'approche qualitative et scientifique inhérente aux domaines de l'édition et de la recherche en sciences humaines et sociales. Le non respect de cette exigence conduit à des erreurs et des manquements (parfois même des contresens) dommageables pour les lecteurs²⁹.

Une interaction forte entre informatique et sciences humaines et sociales doit au contraire permettre de tirer le meilleur parti des outils numériques pour la réalisation d'une édition critique de l'*Encyclopédie*, capable d'offrir des moyens de navigation et d'interrogation innovants, susceptible de susciter de nouveaux types de lecture et de parcours, et de fournir autant d'outils de recherche appropriés au travail des étudiants, des amateurs et des spécialistes.

II. Présentation du projet ENCCRE

Le projet ENCCRE est né d'un constat simple : il n'existe pas d'édition critique de l'*Encyclopédie*, ni au format papier³⁰, ni au format numérique. Et pour cause ! Si l'ouvrage a marqué son siècle et les suivants, les innovations qui ont fait sa force constituent, nous

28. Rappelons que l'*Encyclopédie* concentre plusieurs types d'hétérogénéités dont la concomitance est assez rare : dans les sujets, les façons de les traiter, les auteurs, mais aussi dans le temps car les volumes ne se ressemblent pas, se contredisent parfois, se complètent, se nuancent, mettent en scène des polémiques ou usent de divers subterfuges pour contourner la censure.

29. Tel est par exemple le cas de l'article consacré à la détection informatisée des sources de l'*Encyclopédie* récemment publié par Timothy Allen, Charles Cooney, Stéphane Douard, Russell Horton, Robert Morrissey, Mark Olsen, Glenn Roe et Robert Voyer (« Plundering Philosophers: Identifying Sources of the *Encyclopédie* », *Journal of the Association for History and Computing*, vol. 13, n° 1, printemps 2010). Pour une analyse critique de cet article, voir M. Leca-Tsiomis, « The Use and Abuse of the Digital Humanities in the History of Ideas: how to Study the *Encyclopédie* », *History of European Ideas*, vol. 39, n° 4, 2013, p. 467-476.

30. La Table du Pasteur Mouchon, publiée en 1780, constitue peut-être la seule forme d'édition critique existante permettant de naviguer dans les quelque 74 000 entrées du dictionnaire : voir Pierre Crépel, « Peut-on enfin brûler le pasteur Mouchon ? », *RDE* 31-32, 2002, p. 201-232.

l'avons vu, autant de défis éditoriaux à surmonter. Les nouvelles technologies nous donnent aujourd'hui les moyens d'une telle ambition, et laissent donc enfin la possibilité d'imaginer et de construire ce que ce monument de notre histoire intellectuelle, réalisé « par une société de gens de lettres », mérite assurément que nous lui accordions : une édition fidèle et respectueuse permettant de lui redonner vie, d'en éclairer le contenu, de prendre la mesure du contexte scientifique, philosophique et politique dans lequel il s'inscrit, ainsi que de l'histoire éditoriale complexe et mouvementée qui a été la sienne. Entreprises de longue date, les études sur l'*Encyclopédie* ne sont pas en reste de ce point de vue : qu'il s'agisse des travaux fondateurs de Richard N. Schwab, Walter E. Rex, John Lough, de Jacques Proust³¹, ou du renouveau international des études attesté depuis deux ou trois décennies³², les recherches existent et foisonnent, évoluant de jour en jour dans toute leur diversité, méritant d'être mises à la portée des lecteurs par le biais d'une édition numérique spécifiquement conçue pour les accueillir.

Une telle ambition suppose d'abord une nouvelle méthodologie de travail dont la rigueur éditoriale et scientifique repose nécessairement sur l'établissement d'un dialogue fécond, au sein d'une équipe résolument pluridisciplinaire, entre chercheurs en sciences humaines et sociales, chercheurs en informatique et ingénieurs.

Il s'agit en premier lieu de fonder la conception et, partant, le développement de l'outil électronique sur une politique éditoriale générale élaborée *en amont* par des chercheurs, en relation étroite avec leurs problématiques de recherche : qu'édite-t-on, et comment l'édite-t-on ? Deux questions qui n'ont apparemment rien d'original en ce qu'elles se posent à quiconque entreprend une édition critique, quel qu'en soit le format.

Le changement de support implique néanmoins une réflexion profonde sur le traitement des textes, sous-tendue par une interrogation sur ce qu'est un texte, considéré comme un objet culturel dont il faut à la fois pouvoir respecter l'intégrité et restituer l'inscription dans un contexte susceptible d'en permettre l'interprétation. Si le passage au numérique permet le renouvellement et l'invention de nouveaux modes de lecture, de navigation, d'enrichissement critique ou de recherche, tout n'est donc pas *a priori* souhaitable : les fonctionnalités

31. Voir les travaux cités note 2.

32. Renouveau dont témoignent nombre d'études florissantes actuellement, qu'il s'agisse des revues comme les *Recherches sur Diderot et sur l'Encyclopédie* (Paris) ou le *Recueil d'études sur l'Encyclopédie et les Lumières* (Tokyo), des colloques, des séminaires réguliers et suivis, ou des thèses en cours.

de l'édition doivent être conçues au terme d'une étroite interaction entre les spécialistes de l'*Encyclopédie*, garants de la pertinence et de la scientificité des solutions retenues, et les informaticiens et ingénieurs, capables de mettre au point les technologies requises ou d'en proposer de nouvelles.

Le dernier enjeu concerne la nécessaire évolution des méthodes et pratiques de travail de l'éditeur critique, c'est-à-dire du chercheur en sciences humaines et sociales, dans ce nouveau contexte numérique. Il consiste fondamentalement ici à juguler l'écart susceptible de se creuser entre, d'un côté, la démarche et les compétences d'un chercheur spécialiste d'un corpus particulier et, de l'autre, les technologies numériques déployées par les ingénieurs et les informaticiens pour permettre l'édition électronique de ce corpus. À ce niveau encore, un dialogue est donc requis entre sciences humaines et sociales et informatique, de façon à concevoir des mécanismes (et donc des interfaces) de saisie (et donc d'édition) non seulement directement et facilement utilisables par les chercheurs, mais aussi repensés en fonction de l'ampleur de l'œuvre à éditer et des nouvelles solutions informatiques disponibles.

Politique éditoriale du projet

Avant d'entrer plus avant dans la description du projet, précisons les principes et les objectifs essentiels sur lesquels il repose.

Il s'agit premièrement de fournir une *édition complète*, librement accessible en ligne, qui soit en mesure de donner accès à ce qui constitue la *matérialité de l'ouvrage* : une édition qui permette donc de prendre « numériquement » l'*Encyclopédie* en main, de la feuilleter page après page, d'en consulter simultanément l'original et la transcription, de même que l'ensemble de ses éléments constitutifs (depuis la page de garde jusqu'aux *errata* et privilèges, les textes et illustrations, dans les différentes versions d'une édition donnée), ou encore de restituer les multiples réseaux de circulations qui la sillonnent de part en part (*via* les renvois entre les 74 000 articles et les 2 000 planches qu'elle renferme).

Il s'agit aussi d'une édition *critique* finement articulée avec les études passées et travaux en cours, *fidèle* dans ce qu'elle donne à voir – ce qui implique une séparation claire entre les données encyclopédiques matériellement propres à l'ouvrage (articles, planches) et les données critiques issues des études sur l'*Encyclopédie* –, *justifiée* sous la forme de notes et de notices argumentées, datées et signées par leur rédacteur et, pour finir, *dotée* d'outils de recherche et d'analyse compatibles avec les postulats précédents.

Le projet souhaite en outre donner à lire une édition *ouverte*, c'est-à-dire adaptée à plusieurs types de lecteurs, à l'enseignement comme à la recherche, en proposant plusieurs niveaux de lecture et modes de parcours de l'œuvre, ainsi que des comptes utilisateurs personnels permettant à chacun de mener ses propres investigations.

Le principe suivant concerne le moyen grâce auquel l'équipe d'éditeurs du projet pourra concrètement effectuer le travail de collationnement, de spécification et d'enrichissement critique de l'ouvrage : nous avons ici opté pour un mode de travail *collaboratif*, réalisé par le biais d'une interface de saisie en ligne adaptée. Outre qu'il constitue un gage sérieux de faisabilité d'un point de vue pratique, ce choix innovant répond à un dernier objectif, fondamental à nos yeux : offrir une édition la plus vivante et la plus moderne possible, adossée aux recherches passées et actuelles, et susceptible d'en susciter de nouvelles.

Une interface collaborative d'édition

Comme nous l'avons signalé, l'interface collaborative d'édition de l'*Encyclopédie* disponible sur Wikisource invite sérieusement à réfléchir aux moyens de faire évoluer les pratiques du travail d'édition par le chercheur en sciences humaines et sociales. L'expérience montre indiscutablement à quel point le travail de collationnement et d'enrichissement collaboratif a pu être efficace.

S'il ne nous semble pas souhaitable, pour des raisons scientifiques évidentes, d'ouvrir aussi largement (c'est-à-dire à tout lecteur potentiel) l'interface d'édition, il paraît néanmoins nécessaire d'adopter le principe d'un mode de travail collaboratif similaire entre les futurs éditeurs du projet. Ce choix présente plusieurs avantages essentiels. Il constitue, d'une part, la condition *sine qua non* à la réalisation concrète d'une entreprise éditoriale d'aussi grande envergure. Il permet, d'autre part, de répondre à la diversité des champs de savoirs consignés dans l'*Encyclopédie* par la diversité des chercheurs qui seront impliqués dans le processus d'enrichissement critique de l'édition. Diderot lui-même, évoquant la nécessaire collaboration des savants « spécialistes » à l'*Encyclopédie*, écrivait³³ :

Quand on en vient à considérer la matière immense d'une encyclopédie, la seule chose qu'on aperçoive distinctement, c'est qu'elle ne peut être

33. Diderot, *Œuvres complètes* DPV, t. VII, p. 175-176.

l'ouvrage d'un seul homme [...]. Qui est-ce qui définira exactement le mot conjugué, si ce n'est un géomètre ? le mot conjugaison si ce n'est un grammairien ? le mot azimuth si ce n'est un astronome ? le mot épopée si ce n'est un littérateur ?

De la même façon, l'interface collaborative d'édition du projet permettra à l'historien des mathématiques d'annoter les articles de mathématiques, l'historien de la grammaire les articles de grammaire, etc. Nous pouvons ainsi imaginer une plateforme en ligne partagée, dédiée au collationnement et à l'enrichissement de l'apparat critique, et par là même, un processus d'édition dynamique, envisagé sur le long terme, où viendront se croiser des apports anciens et nouveaux, et se conjuguer des compétences multiples et complémentaires.

Reste néanmoins à concevoir une chaîne de validation éditoriale pensée en conséquence, ainsi qu'une interface adaptée aux pratiques de travail très diverses d'une équipe d'éditeurs dont il faut présupposer, afin de n'exclure personne, qu'ils ne possèdent pas de compétence informatique particulière (si ce n'est la maîtrise des logiciels de traitement de textes les plus courants). Nous ne manquerons pas de revenir sur ces deux points essentiels. Voyons, dans un premier temps, comment l'ensemble de la plateforme d'édition du projet s'organise.

Structure générale de la plateforme d'édition

À cette première interface, collaborative et réservée aux seuls éditeurs, s'ajoute bien sûr une seconde interface : celle de l'édition publiée en ligne, librement consultable, et permettant de lire, de parcourir et d'interroger le contenu de l'*Encyclopédie* (Illustration 3).

Conformément à notre politique éditoriale, cette seconde interface se verra complétée par des comptes individuels sécurisés pour les lecteurs, que chacun pourra créer et où il disposera d'outils de travail personnels : il s'agira par exemple d'outils permettant d'effectuer ses propres annotations, de les classer, de les indexer, d'enregistrer les résultats de certaines requêtes effectuées sur l'interface de lecture, etc. Aux éditeurs, ces espaces utilisateurs seront directement accessibles sur l'interface collaborative d'édition afin de leur permettre aisément de publier (sur l'autre interface) le résultat de leurs travaux personnels.

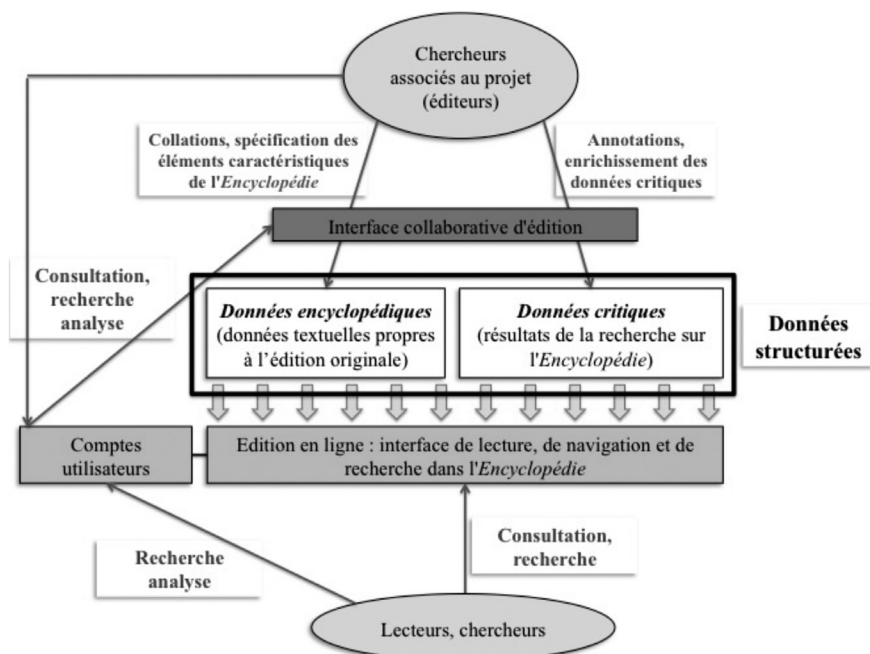


Illustration 3. Structure générale de la plateforme d'édition du projet ENCCRE

Suivant cette logique, les données seront donc produites par le biais de l'interface collaborative, au terme d'une chaîne de validation éditoriale, puis exploitées par l'interface de lecture, de navigation et de recherche pour fournir l'édition accessible au public. Ces données sont de deux types. D'un côté, les *données textuelles* ou *encyclopédiques* (ce qui recouvre aussi bien le texte publié dans les volumes de « discours », les planches que leurs explications) qui correspondent à la version collationnée et spécifiée, conformément à l'original, du contenu de l'*Encyclopédie*. De l'autre, les données critiques qui sont partiellement ou totalement le fruit d'un apport extérieur (celui de l'éditeur, sous forme de notes, de commentaires ou de notices), et supposent donc un certain degré d'interprétation. À quels critères de spécification ces données répondent-elles cependant, et quels types de liens ou de connexions entretiennent-elles les unes avec les autres ? Ces questions, nous l'avons dit, relèvent d'une étape absolument cruciale de notre démarche éditoriale, celle de la structuration des données, dont il convient d'examiner les enjeux de plus près.

Mise au point de la structure de données

Commençons par les données liées au contenu de l'édition originale de l'*Encyclopédie*. L'établissement du texte (entendu au sens large, ce qui inclut donc les planches) implique la définition des modalités d'une transposition minimisant la déperdition d'information entre son support matériel d'origine et le support numérique. Il repose également sur l'élaboration d'une procédure d'encodage approprié qui rende compte de son caractère polysémiotique intrinsèque : un texte n'est pas un simple ensemble de données, mais un assemblage de données *structurées* qui font système entre elles. Il s'agit donc d'abord de se demander quels éléments de l'*Encyclopédie* doivent être décrits, de quelle façon, et comment ils sont reliés entre eux.

Cette première étape d'identification des données encyclopédiques nécessite de définir deux types de structure. La première, dont nous ne donnerons pas le détail, décrit l'organisation matérielle générale de l'*Encyclopédie* en termes d'édition, d'exemplaire, de volumes et de pages : elle correspond à ce que nous pourrions appeler la structure macroscopique des données encyclopédiques. La seconde s'attache, quant à elle, à la spécification du contenu à l'échelle locale, c'est-à-dire à l'échelle d'un article, d'une planche ou de son explication. Voyons plus précisément en quoi elle consiste en prenant pour exemple l'article PORISTIQUE (*Enc.*, XIII, p. 126a) :

PORISTIQUE, adj. (*Mathém.*) quelques auteurs appellent *méthode poristique* la manière de déterminer par quels moyens, & de combien de différentes façons un problème peut être résolu. Voyez PROBLÈME, DÉTERMINÉ, ÉQUATION, RACINE, SOLUTION. *Chambers.* (O).

On peut repérer un certain nombre d'objets particuliers dans cet article : une adresse [PORISTIQUE], une indication grammaticale [adj.], un désignant [(*Mathém.*)], un ensemble de renvois vers d'autres articles de l'*Encyclopédie* [Voyez PROBLÈME, etc.], une mention [*Chambers*] indiquant que le contenu de l'article est issu de la traduction (complète ou partielle) d'un article de la *Cyclopaedia*, ainsi qu'une signature [(O)]. Le texte de l'article peut ainsi être linéairement décrit grâce à un certain nombre de données (les mots) et de métadonnées (les mots possédant un statut particulier). Le principe consiste ensuite à regrouper formellement ces dernières au sein d'un ensemble de *catégories encyclopédiques* considérées comme des métadonnées possédant le même statut, ou pouvant plus généralement être rattachées à un seul et même type d'élément descriptif du contenu d'un article dans l'édi-

tion originale de l'*Encyclopédie*. L'article **PORISTIQUE** sera dès lors finement décrit par l'instance d'une catégorie de type « adresse », d'une catégorie « indication grammaticale », « désignant », « ensemble de renvois », « mention » et « signature », ainsi que d'une dernière catégorie correspondant au « texte par défaut » (c'est-à-dire à l'ensemble des mots de l'article sans spécification particulière). Bien entendu, ces différentes catégories sont liées entre elles par un certain nombre de connexions qu'il convient pareillement de décrire avec la plus grande précision. Les catégories « indication grammaticale », « désignant », « ensemble de renvois », « mention », « signature » et « texte par défaut » entretiendront par exemple un lien d'appartenance à la catégorie « adresse » dans la mesure où les instances de ces différentes catégories décrivent le contenu du texte appartenant à une adresse particulière. De la même façon, l'instance de la catégorie « ensemble de renvois » dans l'article **PORISTIQUE**, parce qu'elle contient elle-même cinq renvois à cinq articles distincts, doit à son tour être décrite grâce à cinq instances d'une autre catégorie encyclopédique : la catégorie « renvoi ». Chacune de ces cinq instances possède un premier lien d'appartenance à l'« ensemble de renvois » (lui-même lié à la catégorie « adresse »), ainsi qu'un second type de connexion extérieure à l'article : celle permettant de le lier avec l'« adresse » ou l'« entrée » de l'*Encyclopédie* à laquelle il nous invite à nous reporter – lorsque ce lien, bien sûr, peut être établi avec certitude ; dans le cas contraire, l'articulation avec des données critiques s'avérera nécessaire. L'instance « **PROBLÈME** » de la catégorie « renvoi » dans l'article **PORISTIQUE** sera donc ici liée à l'entrée « **PROBLÈME** » de l'ouvrage (c'est-à-dire à l'instance « **PROBLÈME** » de la catégorie « entrée » correspondante).

La description de la structure des données des articles de l'*Encyclopédie* conduit ainsi à l'établissement d'une liste de catégories encyclopédiques propres à l'ouvrage ainsi que de la liste des liens qu'elles entretiennent entre elles. La finesse de cette description relève d'un choix éditorial de notre part, effectué en toute connaissance de l'ouvrage, et en toute cohérence avec les objectifs de la politique précédemment exposée : le critère de fidélité à l'original, ou les fonctionnalités de lecture, de navigation et de recherche qui seront proposées sur l'interface de l'édition en ligne, en sont en effet totalement tributaires.

Revenons en particulier sur l'importante question des renvois entre articles. Nous avons précédemment constaté que l'édition d'ARTFL (comme celle fournie par alembert.fr) ne spécifie pas ces éléments en tant que tels : elle se contente au contraire de les identifier par le biais d'un repérage des fragments de textes (hors entrées) typographiés en petites capitales, puis de leur assigner un lien hyper-

textuel vers l'ensemble des résultats de la recherche du mot ou groupe de mots correspondant à ce renvoi parmi l'ensemble des adresses et entrées de l'ouvrage. Cette option éditoriale, qui ne peut être considérée comme une description, ne permet pas d'établir un lien direct, conforme au mode de circulation de l'édition originale, vers l'entrée ou l'adresse spécifiquement indiquée par le renvoi considéré. Elle ne permet pas davantage de traiter le cas des renvois apparaissant sous une forme moins explicite. Penchons-nous par exemple sur l'entrée *JET d'eau* (*Enc.*, VIII, p. 521a-b), où nous nous retrouvons clairement confrontés à cette particularité (Illustration 4). Le renvoi aux articles *GERBE* et *BOUILLONS*, y apparaît en effet de la façon suivante : « Il y a encore des gerbes, des bouillons. Consultez *ces articles* à leur lettre » (*Enc.*, VIII, p. 521a). Pour en rendre compte dans l'édition projetée, il suffit de créer deux nouvelles catégories encyclopédiques, « indication de renvois non explicite » et « renvoi non explicite » dont les instances respectives dans le texte de l'article *JET d'eau* seront liées entre elles, permettant ainsi de naviguer de la première, « Consultez ces articles à leurs lettres », vers les deux renvois non explicites correspondants, « gerbes » et « bouillons », chacun d'entre eux permettant à son tour de se rendre à l'article indiqué (*GERBE* et *BOUILLONS*, respectivement).

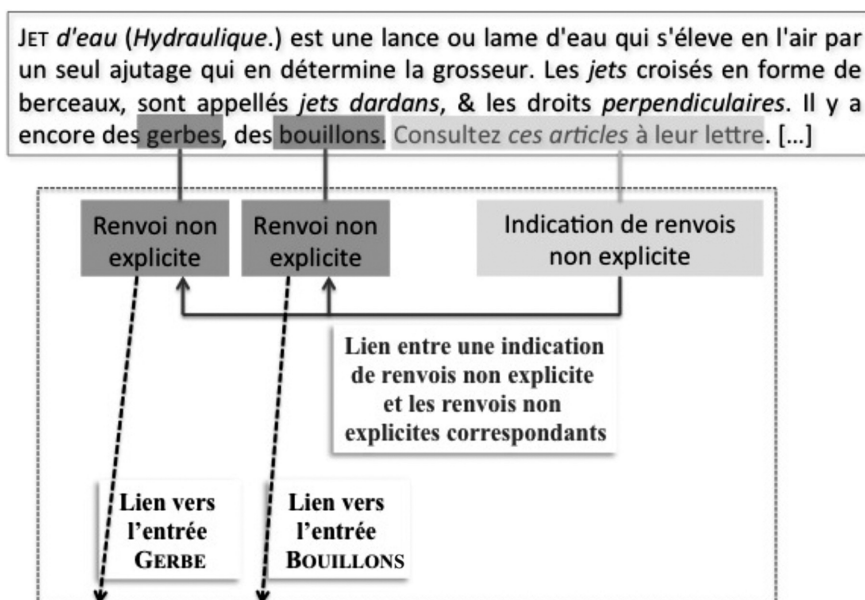


Illustration 4. Représentation d'un extrait de l'article *JET d'eau* et de la solution adoptée, en terme de structure de données, pour modéliser le cas d'un renvoi non explicite dans l'*Encyclopédie*

Suivant le même principe, nous avons plus largement procédé à la création de l'ensemble des catégories encyclopédiques nécessaires à la description du contenu de l'*Encyclopédie*, ainsi qu'à l'établissement d'une typologie précise des liens qui permettront de reconstituer, de façon fidèle à l'original, l'organisation hiérarchique des différents éléments ainsi que les réseaux de circulation correspondants. Il s'agit par exemple des renvois des articles vers les planches, des multiples liens entre les planches (ou parties de ces planches) et leurs explications, des renvois des volumes de planches vers les volumes de textes, de la prise en compte des *errata*, etc. Nous aboutissons ainsi à une structure de données, exclusivement encyclopédiques, qu'il convient désormais d'articuler avec le versant critique de notre édition, c'est-à-dire avec les différentes catégories de données critiques prévues à cet effet.

Cette seconde étape, étroitement dépendante de toutes les études entreprises sur l'*Encyclopédie*, paraît d'autant plus indispensable que la seule structure de données encyclopédiques ne suffit pas à pallier toutes les difficultés éditoriales soulevées par l'ouvrage. Nous avons déjà mentionné le cas problématique des renvois vers des adresses (ou entrées) incertaines ou inexistantes dans l'ouvrage, impliquant un éventuel enrichissement critique du mécanisme de navigation de l'édition, de même que la difficile question des auteurs d'articles. Concentrons-nous, à titre d'exemple, sur ce deuxième problème.

L'ouvrage, comme on sait, contient une myriade d'articles signés par deux collaborateurs ou plus. Tel est le cas de l'article JET d'eau, au sein duquel deux signatures différentes apparaissent : la première « (K) » au terme du premier paragraphe, la seconde « (O) » à la fin du texte.

L'une des premières et des plus légitimes questions que se pose un lecteur moderne est celle de l'auteur, et pour peu qu'il sache que l'*Encyclopédie* est le fruit du travail de nombreux collaborateurs, la seconde est : par quel moyen les reconnaître ? Si, d'après la structure de données encyclopédiques précédemment établie, les deux signatures (K) et (O) de l'article auront préalablement été identifiées comme telles, comment savoir cependant que le (K) désigne d'Argenville et le (O) D'Alembert ? La réponse se trouve, par bribes, dans les textes d'escorte de l'*Encyclopédie*, à des endroits souvent connus des seuls spécialistes (certaines correspondances entre signatures et auteurs sont faites à la fin de l'Avertissement du tome I, d'autres dans le tome III, etc.). La question suivante du lecteur consisterait probablement alors à se demander de quelles parties de l'article d'Argenville et D'Alembert sont respectivement les auteurs, ceci menant tout naturellement à une autre exigence : celle consistant à pouvoir rechercher un terme (ou une

expression) parmi l'ensemble des contributions que les résultats des études sur l'*Encyclopédie* auront permis d'attribuer à un collaborateur particulier dans l'ouvrage. Se posent donc ainsi la question de l'attribution de tel ou tel passage, de tel ou tel article, à tel ou tel collaborateur, ainsi que celle de la justification raisonnée de cette attribution.

JET *d'eau* (*Hydraulique.*) est une lance ou lame d'eau qui s'élève en l'air par un seul ajutage qui en détermine la grosseur. Les *jets* croisés en forme de berceaux, sont appelés *jets dardans*, & les droits *perpendiculaires*. Il y a encore des gerbes, des bouillons. Consultez *ces articles* à leur lettre. (K)

Mariotte démontre qu'un *jet d'eau* ne peut jamais monter aussi haut qu'est l'eau dans son réservoir. En effet, l'eau qui sort d'un ajutage devrait monter naturellement à la hauteur de son réservoir, si la résistance de l'air & les frottemens des tuyaux ne l'en empêchoient. *Voyez l'article* FLUIDE. Mais cette résistance & ces frottemens font que l'eau perd nécessairement une partie de son mouvement, & par conséquent ne remonte pas aussi haut. Ce même auteur a aussi fait voir que lorsqu'un grand *jet* se distribue en un grand nombre d'autres plus petits, le quarré du diametre du principal ajutage doit être proportionnel à la somme de toutes les dépenses de ses branches ; & que si le réservoir a cinquante-deux piés de haut, & l'ajutage six lignes de diametre, celui du conduit doit être de trois pouces. Les différentes regles pour les *jets d'eau* se trouvent renfermées dans un ouvrage exprès de M. Mariotte, imprimé dans le recueil de ses œuvres. *Chambers.* (O)

La solution que notre édition sera capable d'apporter à l'ensemble de ces problématiques éditoriales s'appuie sur plusieurs idées-clés (Illustration 5). Il s'agit d'abord de pouvoir *lier* la signature que le lecteur, novice ou averti, *voit* dans le texte (le (K) de d'Argenville par exemple) à deux autres types d'informations de nature *critique* : le collaborateur (la signature est celle de d'Argenville) et le passage de l'article correspondant (ceci est un paragraphe de l'article attribué à d'Argenville). Il s'agit aussi de faire en sorte que cette liaison entre les trois éléments puisse elle-même faire l'objet des justifications nécessaires par l'éditeur travaillant sur l'article.

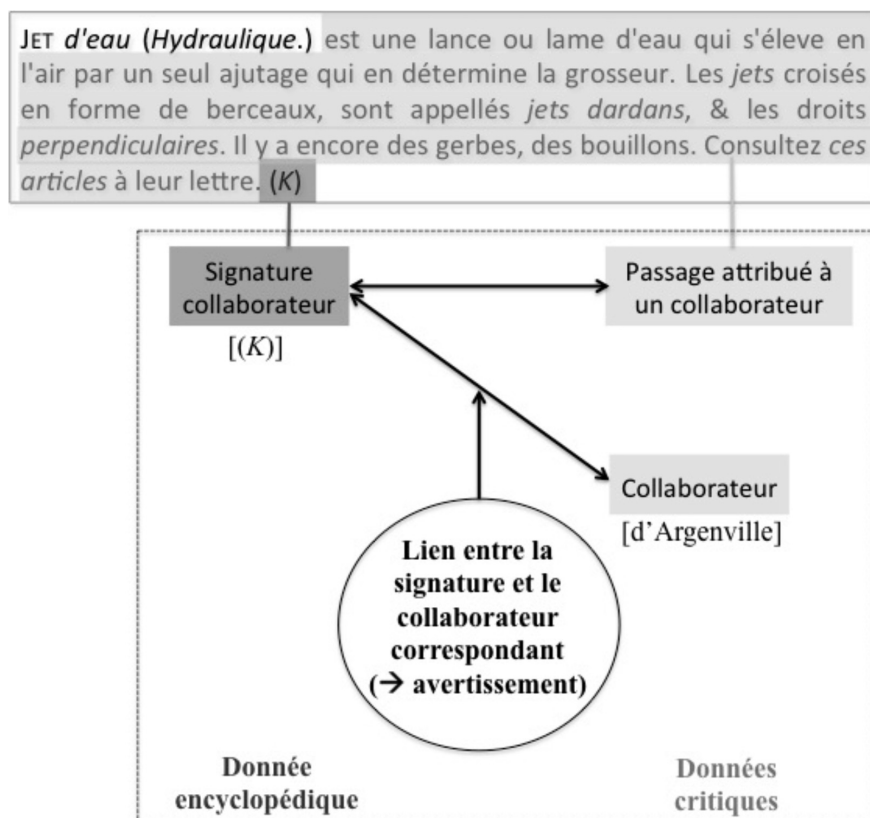


Illustration 5. Représentation d'un extrait de l'article *JET d'eau* et de la solution adoptée, en terme de structure de données, pour modéliser les liens entre la signature (K), le collaborateur d'Argenville, et le passage du texte de l'article pouvant lui être attribué

On voit ainsi se dégager un processus de structuration entre, d'un côté, une catégorie encyclopédique (la signature présente dans le texte) et, de l'autre, deux catégories critiques (le collaborateur correspondant à la signature et l'extrait du texte de l'article auquel cette signature correspond)³⁴. Des liens sont en outre établis pour rendre compte de l'attribution du passage à son auteur, *liens qui doivent pouvoir être annotés*, par exemple en renvoyant à la partie de l'Avertissement qui

34. Ces deux informations critiques n'ont pas le même statut : la première, qui consiste à dire (et à justifier) que « ce signe typographique (K) est à rapprocher de l'auteur d'Argenville, de sa notice etc. », relie une donnée encyclopédique à une donnée extérieure (et donc critique) ; la seconde relie une partie du texte (délimitée par l'éditeur) à la donnée encyclopédique (K).

explique que telle signature désigne tel collaborateur, ou en fournissant une notice bibliographique sur ce dernier, un lien permettant d'accéder à ou rechercher dans l'ensemble de ses contributions dans l'*Encyclopédie*, etc.

Une telle structure, simple dans son principe, permet non seulement d'affiner les futures capacités d'interrogation du moteur de recherche, de prévoir une séparation claire dans l'édition entre ce qui appartient au contenu de l'ouvrage et ce qui relève d'un enrichissement critique de ce contenu (la matérialité de l'original est donc respectée), mais aussi de traiter tout aussi efficacement les autres cas particuliers (nous le savons grâce aux nombreuses études disponibles) susceptibles de se présenter dans l'ouvrage.

L'adresse ALLÉES DE JARDIN (*Enc.*, I, p. 278a-279b) présente par exemple une première partie terminée par la signature (K), suivie d'une seconde partie commençant par *, et donc attribuable à Diderot³⁵. Dans un récent article de la revue³⁶, Fabrice Ferlin a cependant démontré qu'une remarque explicite de D'Alembert dans ses *Opusculs mathématiques* (tome I, 1761) permet sans aucun doute de lui attribuer cette deuxième partie de l'article. Pour traiter un tel cas, et informer le lecteur de ce nouvel apport de la recherche, il suffira de reprendre la structure de données précédente en y apportant les modifications instantielles nécessaires (Illustration 6) : la signature « * » sera certes toujours liée au passage collaborateur correspondant, ainsi qu'à un collaborateur particulier, mais l'instance de cette dernière catégorie (« collaborateur ») correspondra dans ce cas au nom de D'Alembert. Une note sur ce lien, conditionnant en quelque sorte son existence, devra en outre contenir les informations justifiant la correction de cette attribution (ce qui inclut par exemple les références précises à l'article de F. Ferlin, ainsi qu'un lien direct vers sa version en ligne).

De façon similaire, la structure de données s'applique aussi facilement à la multitude des articles non signés dans l'*Encyclopédie* : plus besoin de catégorie « signature » dans ce cas (l'article n'en possédant pas dans l'édition originale), un simple lien annoté (justifiant l'attribution de l'article ou du passage) entre les instances des catégories « passage collaborateur » et « collaborateur » suffira à rendre compte, d'un point de vue éditorial, du processus. Les autres cas de figure s'y adaptent tout aussi aisément, depuis l'erreur typographique³⁷, jusqu'à

35. Signalons que la suite de l'article ne contient plus aucune autre signature.

36. F. Ferlin, « D'Alembert et l'optique », *RDE* 43, 2008, p. 127-144.

37. « On a mis (Z) pour (Q) à l'article *Aide de Camp* » (*Enc.*, I, p. xlvj).

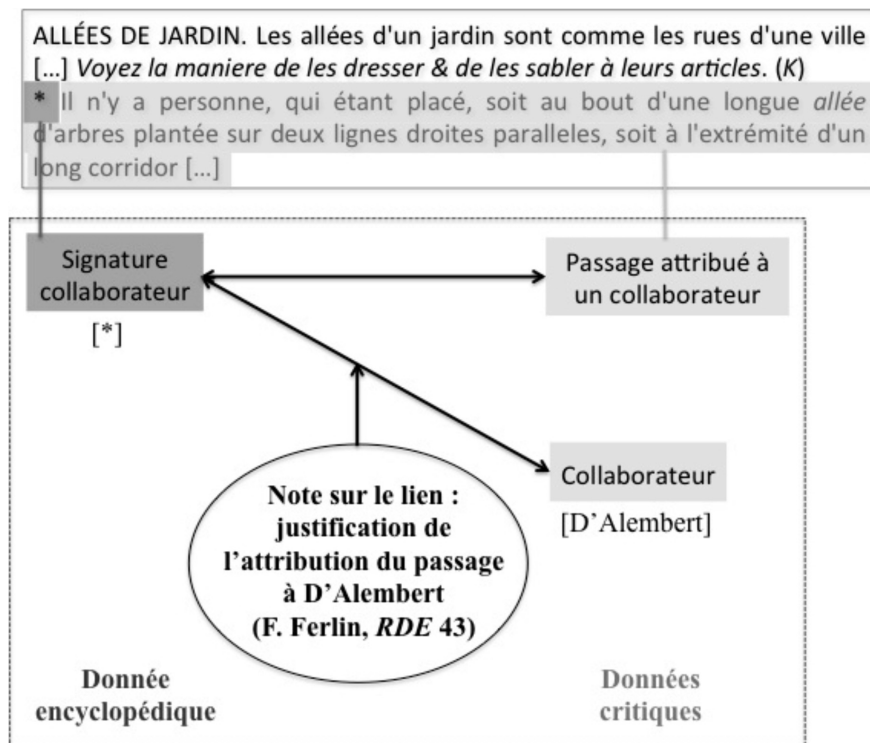


Illustration 6. Représentation d'un extrait de l'article ALLÉES DE JARDIN et de la solution adoptée, en terme de structure de données, pour signaler l'attribution de la partie signée * à D'Alembert (au lieu de Diderot) par le biais d'une note critique

l'absence³⁸ ou l'erreur d'attribution³⁹. Partant de là, de nouvelles possibilités d'interrogation du corpus apparaissent : par exemple, effectuer une collecte ou une recherche dans tous les passages liés à une étoile, ou, plus finement encore, dans tous les passages attribués à Diderot, parce qu'un chercheur dont les travaux font autorité les lui a attribués, ou encore parce qu'une signature * les précède⁴⁰.

38. Voir par exemple les arguments invoqués par John Lough dans *Essays on the Encyclopédie of Diderot and D'Alembert*, ouvr. cité, note 2, chapitre « D'Holbach's contribution ».

39. Les *errata* du tome V signalent, par exemple, une erreur d'attribution : « Consultation, (*Méd.*) pag. 109 ; col. 1, lig. 39. Article de M. Bouillet fils, lis. de M. d'Aumont. »

40. L'éditeur Diderot se reconnaît parfois par une étoile *au début* de l'article, parfois par l'*absence* de marque : les encyclopédistes indiquent que si « plusieurs articles appartenant à la même matière », « immédiatement consécutifs », sont « faits

Des possibilités éditoriales offertes par cette structure de données

Ce travail de structuration des données encyclopédiques et critiques a bien sûr été étendu à toutes les caractéristiques éditoriales de l'ouvrage en tenant compte de l'ensemble des problématiques de recherche explorées dans la très riche littérature existante⁴¹. Nous sommes ainsi parvenus provisoirement à une liste de 59 catégories encyclopédiques et critiques, liste complétée par une table énumérant les 191 liens possibles entre ces différentes catégories et détaillant, pour chacun d'eux, les fonctionnalités de lecture et de navigation qui y seront associées.

En précisant clairement en amont ce qui est d'ordre textuel et ce qui est d'ordre interprétatif, ainsi que les relations que ces métadonnées entretiennent entre elles, l'étape de structuration des données garantit les conditions d'élaboration d'une interface d'édition permettant d'éviter au lecteur toute confusion entre ce qui relève respectivement de l'*Encyclopédie* originale et des recherches dont elle a fait l'objet.

La conception novatrice du *lien informé (ou annoté)* qui sous-tend notre démarche permet en outre d'envisager un processus d'annotation critique de l'*Encyclopédie* à plusieurs échelles, tirant ainsi le meilleur parti des possibilités offertes par le format numérique. Les éléments constitutifs de l'*Encyclopédie* (textes d'escorte, articles, sections de planches ainsi que leurs explications) pourront « classiquement » faire l'objet d'éclaircissements critiques qui prendront la forme d'introductions, de présentations, de commentaires portant sur l'ensemble de ces unités significatives, ou celle de notes ponctuelles portant sur des unités de moindre ampleur (une partie de l'un de ces ensembles, voire tel terme méritant éclaircissement). Les grandes pages in-folio des planches seront elles-mêmes (informatiquement) subdivisables à l'instar des « passages » de texte, et pourront donc pareillement être annotées de façon globale ou plus ponctuelle (sur

ou revûs » par « la même personne », la désignation n'apparaît qu'« à la fin du dernier de ces articles » (*Enc.*, I, p. lxxv).

41. La structure de donnée obtenue a par ailleurs été enrichie de la typologie des métadonnées établie par l'équipe japonaise du professeur Yoichi Sumi de l'Université Keio de Tokyo, qui forme l'un des principaux groupes de chercheurs participant au projet ENCCRE. L'édition permettra donc d'intégrer le résultat du minutieux travail de repérage de métadonnées que cette équipe a réalisé à l'échelle de plusieurs volumes de l'*Encyclopédie*. Voir l'exposé de ces travaux dans la revue *REEL, Recueil d'études sur l'Encyclopédie et les Lumières*, ISSN 2186-4284, n° 1, mars 2012 ; voir aussi Yoichi Sumi et Takeshi Koseki, « Pour une édition critique informatisée de l'*Encyclopédie* : quelques précisions sur les métadonnées », *RDE* 44, 2009, p. 207-218 ; Yoichi Sumi, « Puiser aux sources de l'*Encyclopédie* », *RDE* 46, 2011, p. 227-231.

telle ou telle partie de l'image, découparable à volonté) : ceci constitue assurément un apport nouveau par rapport aux éditions critiques existantes (numériques ou au format papier), qui concerne une partie de l'*Encyclopédie* trop souvent oubliée, et pourtant si riche et complexe⁴². La dernière innovation du processus d'annotation critique consiste en la possibilité d'annoter et de former des liens entre différents types d'objets, encyclopédiques et/ou critiques. Elle repose sur un modèle complexe de structuration et d'exploitation des données dans lequel les liens (hypertextuels par exemple) deviennent davantage que de simples associations entre deux objets informationnels : ils sont à proprement parler porteurs d'informations, dont l'explicitation, l'explication ou la justification prennent la forme de notes éditoriales.

Ce concept de lien informé et la notion de catégorie sur lesquels s'appuie notre approche offrent le précieux avantage d'être très facilement adaptables à d'autres types de corpus possédant un niveau de complexité plus simple ou comparable à celui de l'*Encyclopédie*. Les catégories textuelles et critiques ayant préalablement été définies comme des classes abstraites (indépendantes, donc, d'un corpus particulier), il suffira en effet de les re-décrire en fonction des caractéristiques de l'œuvre à éditer et de dresser la typologie des liens correspondante afin de parvenir à une nouvelle structure de données spécifiquement adaptée à cette œuvre. La plateforme d'édition que nous projetons de développer sera donc théoriquement transposable à d'autres dictionnaires et encyclopédies de l'époque, de même, par exemple, qu'aux volumes d'*Histoire et Mémoire de l'Académie des sciences*, laissant ainsi imaginer pouvoir déplacer l'outil d'un point à l'autre du phénomène complexe de circulation des savoirs qui caractérise le mode de fabrication d'œuvres lexicographiques au XVIII^e siècle.

Protocole et méthodologie de travail

Dans la continuité du travail effectué sur la structuration des données se pose bien sûr aussi la question de la faisabilité pratique de cette vaste entreprise. L'un des deux principaux défis à relever de ce point de vue sera sans nul doute celui de l'interfaçage ou, plus précisément, des diverses interfaces par le moyen desquelles la chaîne

42. Surtout lorsqu'on se rappelle les nombreuses modifications, liées aux délais de publication des différents volumes, dont elle a fait l'objet, ou le changement intervenu dans la conception même de la Description des Arts : voir R.N. Schwab, *Inventory of Diderot's Encyclopédie*, vol. VII, Oxford, 1984, p. 5 à 16, J. Proust, *Diderot et l'Encyclopédie*, ouvr. cité, note 2, p. 207, ainsi que différentes études de Madeleine Pinault-Sørensen.

éditoriale numérique complète pourra être assurée. Il s'agit d'abord de l'interface d'édition, qui devra permettre aux chercheurs sur l'*Encyclopédie* de toutes origines disciplinaires de réaliser efficacement et commodément (sans codage) leur travail d'éditeur. Il s'agit ensuite de l'interface de lecture, de navigation et de recherche consultable par les internautes. Il s'agit enfin, au confluent des deux, d'offrir un certain nombre d'outils de recherche et d'analyse, rassemblés dans des comptes utilisateurs personnels mis à la disposition de tout lecteur ou de tout éditeur.

Nous avons déjà eu l'occasion de souligner le rôle crucial de l'interface collaborative d'édition dans notre projet. C'est à l'intérieur de cet espace accessible en ligne que l'éditeur spécifiera en effet les différents éléments caractéristiques du sous-ensemble de l'*Encyclopédie* dont il aura la charge, sur la base des catégories encyclopédiques prédéfinies (ce qui reviendra par exemple à déclarer tel groupe de termes comme une adresse, une entrée, ou une signature de la même façon que l'on fait passer un passage en italique, en gras ou en petites capitales sur un logiciel de traitement de texte). C'est aussi dans cet espace que ses notes (sur le texte ou sur des liens), ses commentaires et ses notices critiques pourront être saisies avant publication sur l'interface de lecture, de navigation et de recherche. La réussite du projet dépend donc tout particulièrement de notre capacité à concevoir et à développer une interface intuitive et adaptée au vaste champ de compétences et de méthodes de travail de notre équipe d'éditeurs. Pour y parvenir, une étroite interaction entre ingénieurs et chercheurs s'avère indispensable. Elle reposera sur une méthode simple et efficace : les premiers, chargés du développement de l'interface, proposeront régulièrement des prototypes ; les seconds les testeront, rendront compte de leurs faiblesses pour que celles-ci soient corrigées dans la mouture suivante, et ainsi de suite, jusqu'à l'obtention d'une interface exploitable.

Cet impératif de dialogue s'applique bien sûr aux deux autres versants du travail d'interface : celui de l'édition accessible et interrogeable en ligne, et celui des outils d'analyse et de recherche proposés dans les comptes personnels. Ici encore, des tests (par un public de tous horizons, amateur et spécialisé) seront nécessaires, afin de parvenir à une interface de lecture, de navigation et de recherche à la fois intuitive, ergonomique et strictement compatible avec le principe de séparation claire entre ce qui relève de l'œuvre originale et de l'apport critique.

Le projet d'édition que nous avons conçu pose en outre des questions informatiques substantielles, liées pour une part au modèle complexe de structuration des données sur lequel il s'appuie, pour une autre part aux fonctionnalités d'annotation sur les planches prévues dans notre modèle et, pour une dernière part, à l'exigence de généricité

du module central d'édition sur lequel les fonctionnalités des différentes interfaces prendront appui. De ce dernier point de vue, la perspective éditoriale est double : il s'agit à la fois de construire un module capable de s'adapter à des modifications de la structure de données établie (il n'est pas à exclure que cette structure puisse en effet être ultérieurement retouchée en fonction des dernières avancées des recherches sur l'*Encyclopédie*), et, nous l'avons signalé, capable de prendre en charge d'autres types de corpus répondant à un degré de complexité similaire. Le projet ENCCRE se développe donc actuellement en collaboration avec des informaticiens de l'INSA de Lyon (LIRIS) et d'autres projets d'édition numérique (Archives Desanti, Projet Stendhal, Projet des « Dossiers de *Bouvard et Pécuchet* de Flaubert ») autour d'« une approche générique pour la construction d'éditions électroniques de corpus documentaires multistrukturés ». Une allocation doctorale de recherche en informatique, dont le sujet porte sur ce problème spécifique, vient d'être obtenue, et commencera en octobre 2013 pour une durée de trois ans.

La conception et le développement des deux interfaces ainsi que des autres « briques » informatiques du projet seront réalisés par des ingénieurs au cours du même laps de temps, en étroite collaboration avec l'équipe d'informaticiens pour ce qui relève de la compatibilité avec le module central de l'outil d'édition, et avec l'équipe de chercheurs-éditeurs du projet pour ce qui a trait aux phases de tests des prototypes successifs.

Le premier résultat du projet ENCCRE sera livré en ligne dans un délai de dix-huit mois environ. Il correspondra à une première version de l'interface de lecture, de navigation et de recherche, alimentée grâce à l'enrichissement de la saisie produite sur le Wikisource dédié à l'*Encyclopédie*. Les fonctionnalités d'accès aux données critiques seront implémentées au fur et à mesure des dix-huit mois suivants. L'interface collaborative, plus étroitement dépendante du module central d'édition de la plateforme, sera quant à elle ouverte d'ici trois ans. Le travail collaboratif de l'équipe d'éditeurs critiques pourra alors débuter, pour ne plus jamais s'interrompre.

Dans cette perspective, la constitution d'un groupe de chercheurs aux compétences complémentaires s'avère bien sûr primordiale. Soixante ans de recherches depuis les travaux fondateurs de Jacques Proust et de John Lough, puis de Richard N. Schwab, ont cependant permis « d'entrer dans la forteresse⁴³ », en évitant les contresens, en

43. Piste de travail essentielle lancée par J. Proust, « Questions sur l'*Encyclopédie* », *RHLF*, janvier-février 1972, p. 52.

multipliant les approches méthodologiques. Une équipe internationale d'une trentaine de chercheurs est déjà réunie. Citons d'abord l'équipe du professeur Yoichi Sumi de l'Université Keio de Tokyo et d'autres universités japonaises, largement engagée depuis plusieurs années dans le travail de repérage des métadonnées de l'*Encyclopédie*⁴⁴, ainsi que de nombreux autres chercheurs appartenant à divers centres de recherches scientifiques : l'Institut de mathématiques de Jussieu (CNRS, UPMC, Université Paris Diderot), les universités de Lausanne, de Paris-Ouest-Nanterre, le laboratoire LIRE (Universités Lyon 2, Grenoble 3, Saint-Étienne, ENS de Lyon), l'IRCL (Université Montpellier 3), le SYRTE (Observatoire de Paris), l'Institut universitaire européen de Florence, etc. Cette équipe s'appuie par ailleurs sur le Groupe d'édition critique des *Œuvres complètes* de D'Alembert, soutenu par le Comité D'Alembert de l'Académie des sciences, ainsi que sur la riche expérience régulièrement accumulée, depuis 1986, dans la présente revue. Elle reste enfin chaleureusement ouverte à tous les spécialistes désireux de la rejoindre et d'apporter leur contribution à ce projet⁴⁵.

Alexandre GUILBAUD (UPMC), Irène PASSERON (IMJ),
Marie LECA-TSIOMIS (U. Nanterre), Olivier FERRET (U. Lyon 2),
Vincent BARRELLON (LIRIS) et Yoichi SUMI (U. Keio)⁴⁶

44. Voir la note 41.

45. Pour prendre contact avec l'équipe du projet ENCCRE, écrire à : projet.enccre@gmail.com.

46. Sont également signataires : Tatsuo Hemmi (U. Niigata), Takeshi Koseki (U. Hitotsubashi), Hisashi Ida (U. Aymo-Gakuin).